iCoSys

Institute of Complex Systems
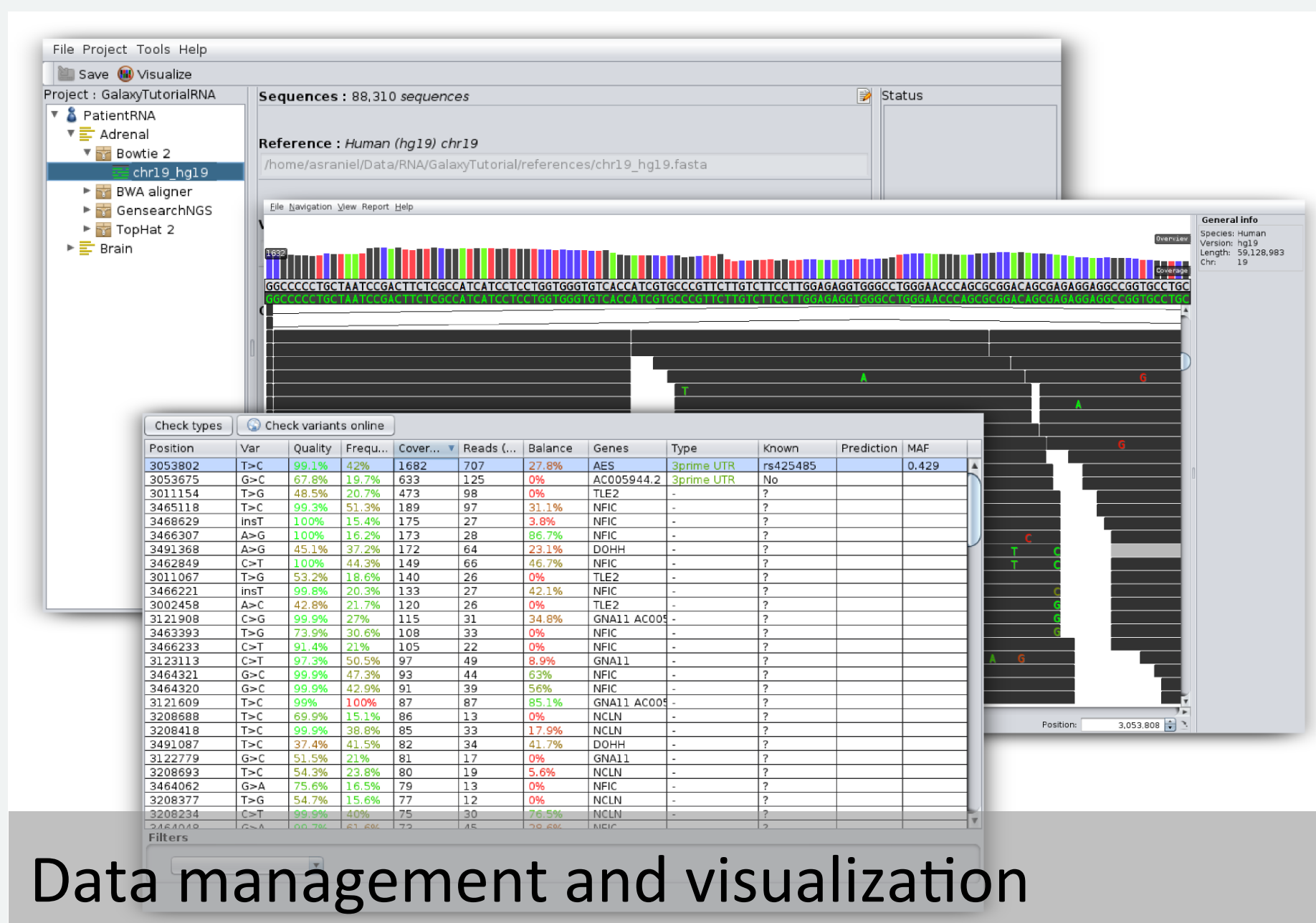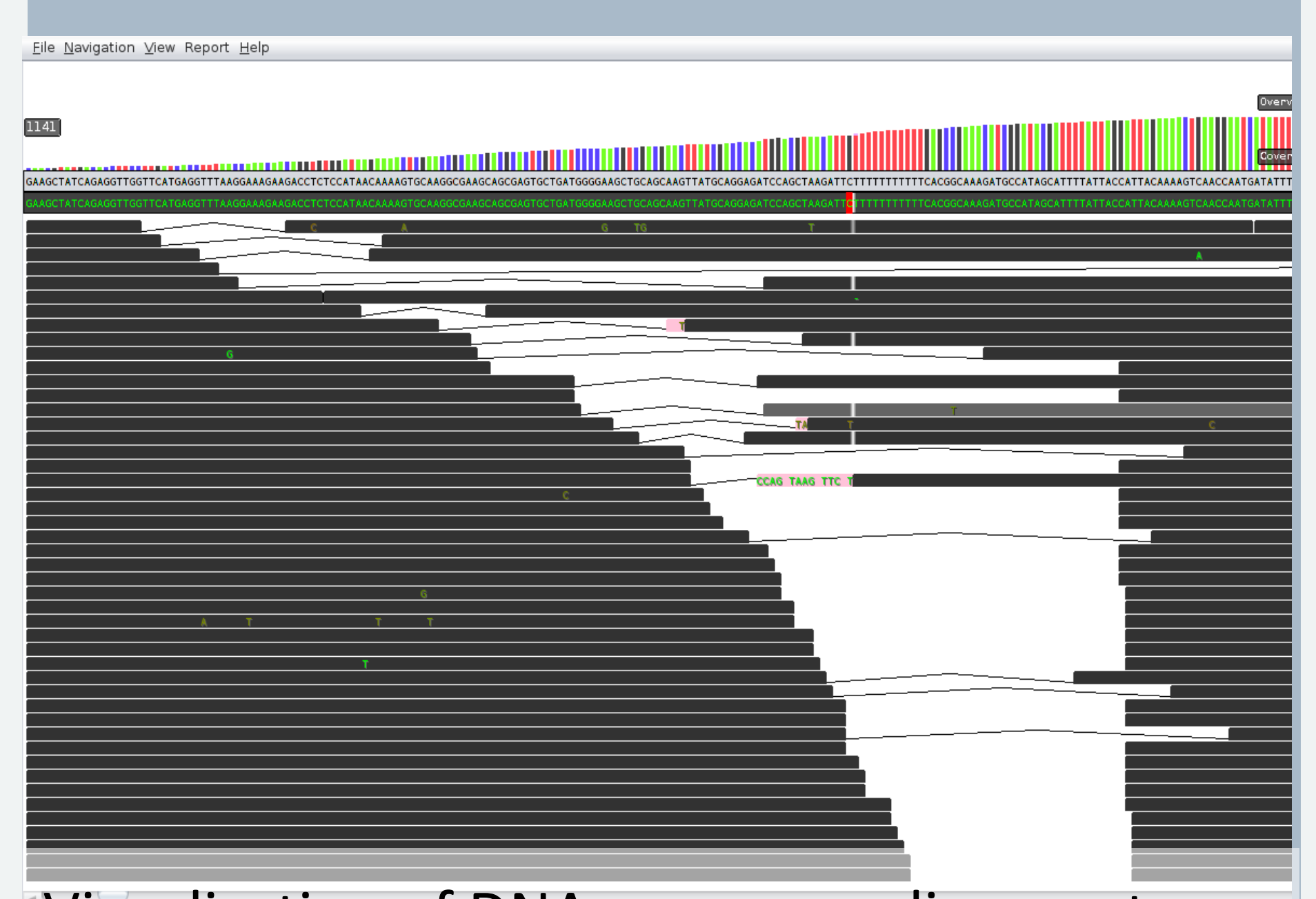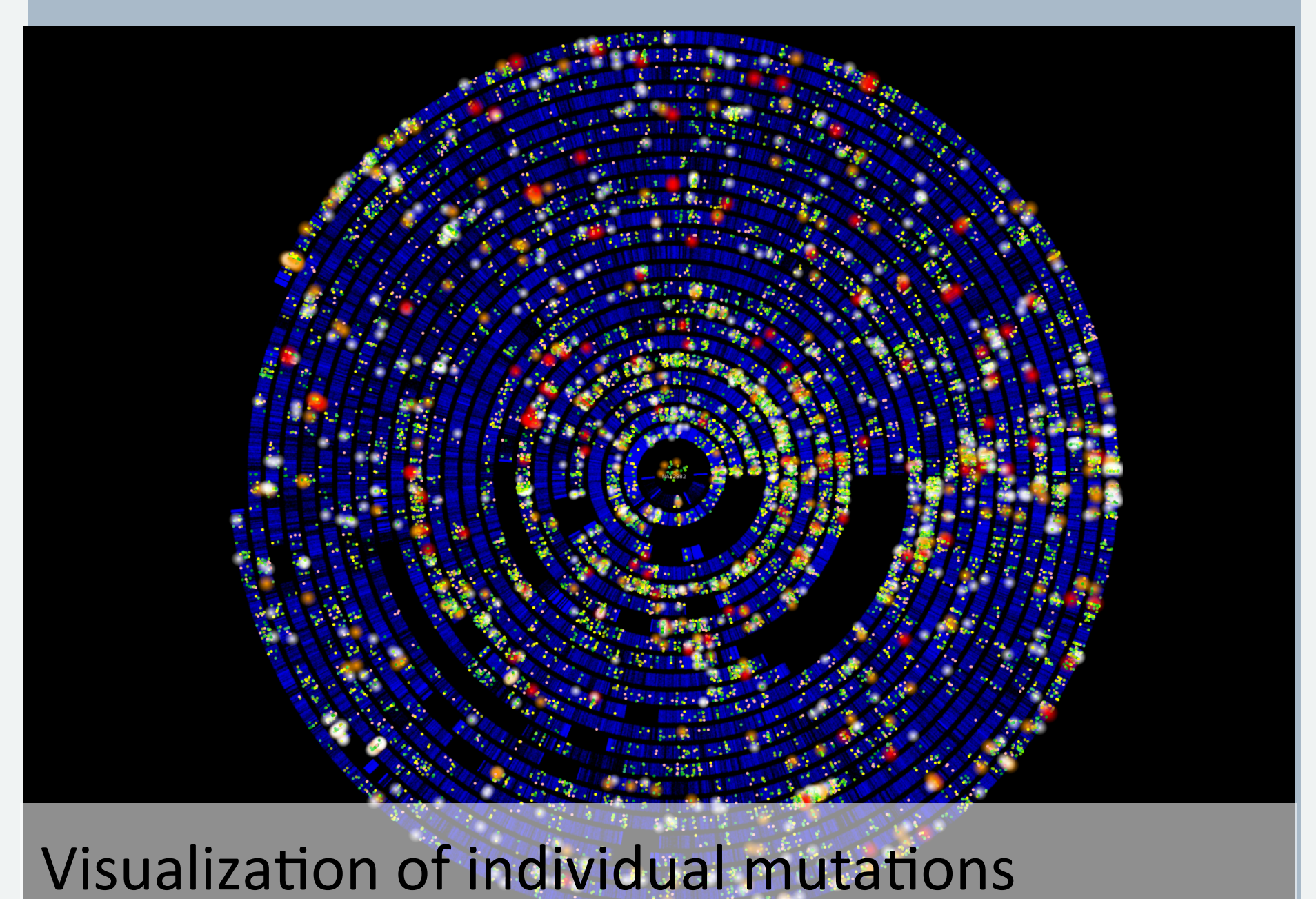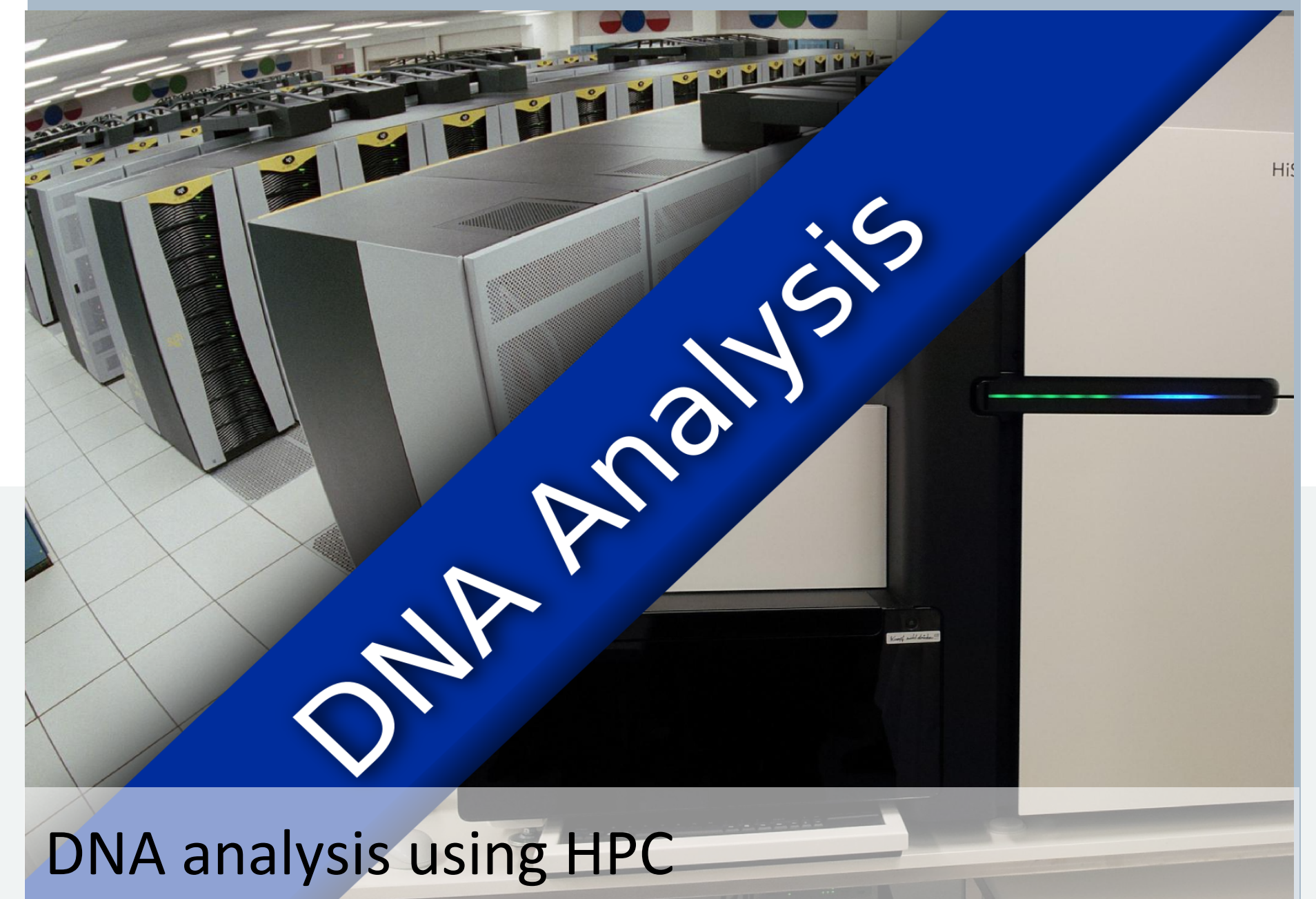

DNA analysis using HPC

DNA Analysis

Context

The study of DNA is a recent and fast evolving field. While the DNA helix has been discovered in 1953 by Watson and Crick, it only became possible to digitalize DNA of living organisms over the past 20. Projects like the human genome project gave the DNA sequencing technologies a great boost, making it possible for the DNA sequence of every organism to be extracted in digital form. The advancements in DNA sequencing technologies have increased the speed at which data is generated at a faster pace than computing technology has evolved, creating a gap between the speed at which data can be produced and at which data can be processed. To handle this challenge of huge amounts of data created during the sequencing process, innovative solutions in various fields of computer science are needed, creating synergies between different fields of research. Storage capacity, processing power, big data and efficient visualization are only a few of the problems to be solved, using technologies like distributed computing and modern visualization techniques. With its focus on complex systems, iCoSys approaches the different challenges with its expertise in various techniques acquired in the different domains of computer science.

Trisomic gene analysis

Data management and visualization

Data analysis

Visualization of individual mutations
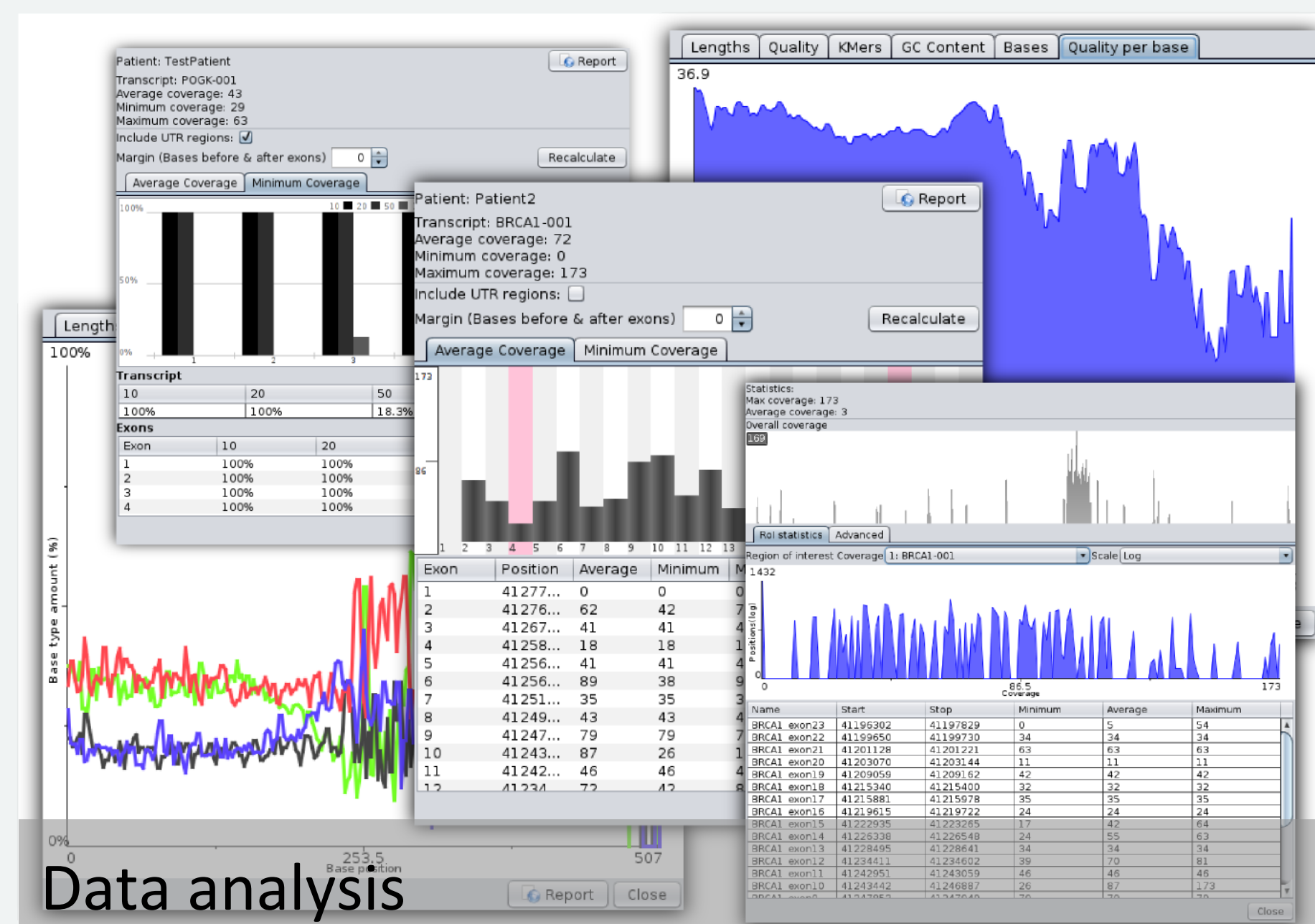
Results

Through a collaboration with the SME Phenosystems SA and the University of Würzburg in Germany, different projects have been realized in the domain of DNA analysis. Those projects include the development of the tool GensearchNGS, as part of a PhD thesis. This tools allows users to diagnose genetic diseases through an user friendly interface, using "Next generation sequencing" data. The focus of this project was on the integration of different tools and algorithms to allow for an easy analysis of DNA sequence data, targeted at small and medium laboratories. The knowledge gained through the GensearchNGS project allowed for a variety of research projects to be launched, exploring different aspects of the challenges raised by the advancement of DNA sequencing technologies. Those projects explore the possibilities of distributed processing using various technologies, including the programming languages POP-Java and POP-C++ developed in house, allowing for easy development of distributed applications in heterogeneous computing environments including grids, clouds and traditional clusters.

Visualization of DNA sequence alignments

Research domain
Bioinformatics, distributed computing,
Data analysis/visualization, HPC

Ecole d'ingénieurs et d'architectes de Fribourg
Hochschule für Technik und Architektur Freiburg

Hes·so// FRIBOURG FREIBURG
Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts
Western Switzerland

PhenoSystems
GensearchNGS
MOLECULAR GENETICS MADE EASY

POP Java

BIOZENTRUM
UNIVERSITAT
WÜRZBURG